

Exploiting the Metathesaurus Update Model

Nels E. Olson, Mark S. Erlbaum, M.D., Mark S. Tuttle, David D. Sherertz,
Octavio Suarez-Munist, Stephanie S. Lipow, William G. Cole, Ph.D.

Lexical Technology, Inc., Alameda, California

Stuart J. Nelson, M.D., National Library of Medicine, Bethesda, Maryland

Five Metathesaurus change files accompany the 1996 NLM UMLS Knowledge Sources.¹ These files are designed to help developers update applications built using the 1995 Metathesaurus. Additionally, they serve as a model of how changes appearing in a new version of a terminology can be explained to those using the previous version. The problem of detecting, interpreting, and implementing changes in controlled medical vocabularies has been described by Cimino² and commented on by Tuttle and Nelson.³

As indicated below, three of the five change files enumerate the concepts (CUIs), terms (LUIs), and strings (SUIs), respectively, that were deleted in the 1996 version; the other two files enumerate the concepts and terms, respectively, that merged with other concepts and terms. Each row in each file represents a deletion or merge. The percentage change in the 1995 version is shown in the last column:

Change File	Row Count	% Change
DELETED.CUI	7,797	3.50%
DELETED.LUI	8,815	2.68%
DELETED.SUI	11,982	2.71%
MERGED.CUI	3,429	1.50%
MERGED.LUI	104	0.03%
TOTALS	32,127	3.23%

Depending on the way that a given application attaches function or content to Metathesaurus CUIs, LUIs, or SUIs, one or more of these files may have to be examined manually.

DELETED.CUI file. Typically, two forces drive the deletion of concepts in the Metathesaurus: obsolescence occurs when a concept ceases to have relevance to the medical domain; evolution occurs when science calls for new nomenclature, e.g., the renaming of a disease. Both have important (but different) ramifications for developers. The specific reasons why concepts (as well as strings) disappear from the Metathesaurus are the result of these forces. They include determinations that: (1) the concept is somehow formally incorrect; (2) the source in which it was originally contained has deleted it in its latest edition; or (3) knowledge has

evolved so as to supersede it or to make it entirely obsolete. Sometimes a concept is merged into another pre-existing concept. In this case the merged concept is not considered to be deleted even though its CUI will no longer be present; it is listed instead as having been merged in the MERGED.CUI file, as described below.

DELETED.LUI file. The record of a deletion of a term in this file does not necessarily imply the appearance of the corresponding concept in the DELETED.CUI file.

DELETED.SUI file. Our computational analysis suggests that many strings have been deleted because they present only trivial differences with various preserved variants of the Preferred Form, e.g., the now deleted "S0011887|AIDS related lymphoma", for which the Preferred Form is: "S0058923|Lymphoma, AIDS-related".

MERGED.CUI files. These represent instances of synonymy that were missed in the previous version of the Metathesaurus. Rows in this file denote synonymy either detected by a human editor or brought to light because of synonymous relationships that have been included as part of newly added Metathesaurus terminologies.

MERGED.LUI files. These merges can come only from changes to the definition of lexical variance.

While the change files represent an exact specification for updating, it can be frustrating not to know why changes were made. An improved specification would include the reason for each change.

1. UMLS Knowledge Sources, 7th Experimental Edition, January 1996, Unified Medical Language System, U.S. Department of Health and Human Services, National Institutes of Health, National Library of Medicine.
2. Cimino JJ. Formal Descriptions and Adaptive Mechanisms for Changes in Controlled Medical Vocabularies. *Meth Inform Med* (to appear).
3. Tuttle MS, Nelson SJ. Editorial Commentary — A Poor Precedent. *Meth Inform Med* (to appear).